Reprinted From The Journal of The American Medical Association
February 1, 1964, Vol. 187, pp. 352-355

Evaluation of Toxicity in Long-Term Clinical Trials

Edward D. Freis, MD, Washington, DC

ALL EFFECTIVE DRUGS are potentially toxic. The toxicity of a new agent is relative and must be evaluated in relation to the drug's effectiveness as well as to the severity of the disease being treated. If the disease is serious, an increased incidence of toxicity can be tolerated in a new drug if it is more effective than previously known agents. On the other hand, if the disease is minor then only minimal toxicity will be tolerated regardless of how effective the drug may be. The three factors—toxicity, therapeutic effectiveness, and the severity of the disease—must be considered in relation to each other.

Preliminary Procedures

A sequence of routine steps should be taken before the submission of a large series of patients to a long-term trial of a new agent. The investigator must be supplied by the sponsor with the chemical structure and animal pharmacology of the drug. The acute minimal lethal dose by various routes of administration in several species of laboratory animals should be available. The investigator must be supplied with chronic toxicity data including tests of cardiovascular, hepatic, renal and bone marrow functions, body weight, and behavioral changes in animals and in the litters as well as pathological studies including histological examination of the organs.

Clinical pharmacological data must be supplied or determined by the investigator prior to the undertaking of long-term studies. The time-dose relationship (times of onset, peak, and disappearance of effect) should be established in man preferably by multiple routes of administration. The investigator must know the duration of action of the drug and its general effects, both favorable and adverse, at various dosage levels in man before he can intelligently plan a dosage schedule. If the acute trials are encouraging, a limited number of patients can be subjected to continuous treatment for a period of three to six months under the direction of a single qualified clinical pharmacologist, preferably the investigator who carried out the short-term trial. The investigator will be less inclined to publish prematurely if he is not competing with others for priority. If severe toxicity does not occur in the small series, the long-term trial can be extended and additional investigators enlisted in the study. These various steps represent a routine sequence that has been recognized as good practice for many years.

Experimental Design

There is a common misapprehension that new drug evaluation depends primarily on the treatment of large numbers of patients. When a large-scale study is begun prematurely there is the unnecessary risk of exposing many patients to unknown toxicity. Large numbers will not provide accurate evaluation if the studies are badly designed. A poorly conceived and executed clinical trial is worse than none since the profession will be misinformed as to the true merits and demerits of the drug.

If the new agent appears to be safe and effective after three or four months of treatment in a limited number of patients, a well-designed double-blind study should then be undertaken utilizing larger numbers of patients. The double-blind crossover technique usually is advisable, in which case each patient is treated for a certain number of months with active drug and then for a similar period with placebos. The sequence of allocation of patients to active drugs or placebos is randomized. A slightly more elaborate, but often more valuable, crossover experiment is the utilization of three regimens—the drug under test, a placebo, and another medication generally recognized as the most satisfactory therapy available to date. In some instances, such as in the treatment of life-threatening diseases, placebo control is not advisable; a comparison is then made between the double-blind method and one using a known effective medication. Certain types of agents, such as anticholinergic compounds, can be identified by noting the characteristic side effects and they must be compared with a drug producing similar subjective reactions.

Dr. Freis is medical investigator at the VA Hospital and associate professor of medicine at Georgetown University School of Medicine.
Presented as part of a Symposium on Safety Evaluation of New Drugs before the Joint Meeting of the Society of Toxicology and the Section on Experimental Medicine and Therapeutics at the 112th Annual Meeting of the American Medical Association, Atlantic City, NJ, June 17, 1963.

The important decision to make about a new drug is not whether it is moderately effective or reasonably nontoxic but rather whether it possesses certain advantages over already existing medications. The literature contains many glowing reports of new agents that in time are found to be inferior to long-established preparations. A better estimate of the value of such new drugs could have been made if the investigators had included a controlled comparison with a preparation of known effectiveness and toxicity.

The double-blind technique is useful for determining the incidence of subjective side effects. The uncritical investigator may be seriously misled about the incidence of subjective side effects when he fails to make a double-blind comparison with a placebo-treated control group. Reports appear frequently in which a high incidence of fatigue, anxiety, headache, weakness, and similar "side effects" are ascribed to a test drug by the investigator who conscientiously records the patient's every complaint. Experience has shown that placebo-treated controls often volunteer similar complaints. Failure to make comparisons with a placebo control group often leads to overestimation of side effects.

The double-blind technique is useful not only in evaluating the incidence of side reactions but is valuable as well in judging drug effectiveness. It is surprising how high the incidence of subjective effects of all types can be in a placebo-treated population. This high rate applies to relief of symptoms as well as to the incidence of side effects. In certain chronic symptomatic conditions subjective improvement may occur in as high as 50% of patients on placebos alone. The falsely optimistic estimate of drug effectiveness is not limited to subjective responses. Even such "objective" measurements as the casual blood pressure are highly conditioned by the subjective mood of the patient and can be influenced by placebos.

Detection and Evaluation of Severe Toxicity

The target organs affected by toxic drugs usually are the liver, kidneys, and bone marrow. Sensitivity reactions such as dermatitis and arthritis also are common. The investigator must be familiar with the chronic toxicity data in animals supplied by the sponsor. Routine laboratory procedures such as serum glutamic oxaloacetic transaminase determination, sulfobromophthalein excretion, urinalysis, blood urea nitrogen determination, and blood count should be carried out before treatment and at regular intervals during treatment.

Unfortunately, many serious toxic reactions cannot be anticipated. In the author's experience the suicidal depressions associated with administration of reserpine or the color-blindness occurring with administration of certain amine oxidase inhibitors could not have been discovered by the usual animal investigations or by routine laboratory tests in patients. Rauwolfia serpentina was used for many years in India and the drug was dispensed widely in this country before its association with severe mental depression became recognized. Routine laboratory procedures cannot substitute for the alert, unhurried investigator who critically reviews the patient's complaints.

When a patient being treated with a new drug develops a severe complication the investigator may not know whether the reaction occurs because of the drug or because of some chance factor. If the complication occurs in several patients he can be reasonably certain that it is caused by the drug; but, when it occurs only in an isolated case he will have difficulty in relating the reaction to drug effect. The appearance of a single case of albuminuria or of thrombocytopenic purpura may or may not be related to the compound under study. In such patients the new agent should be discontinued immediately and the patient should be hospitalized. Pertinent laboratory tests, including the taking of a biopsy specimen of the affected organ when indicated, may be helpful in deciding whether the complication was drug-induced. It is often necessary to reinstitute a small challenging dose of the drug after the patient has recovered fully and is still in the hospital. If no adverse effects appear in the functional tests following this challenge, the individual doses and frequency of administration may be gradually increased. The dosage level which existed prior to the complication is thus approached while the patient is observed closely for signs of recurrence. If the patient remains free of a toxic manifestation he can be discharged on the original dosage schedule. However, he must return at frequent intervals for continued clinical and laboratory check-ups since the toxic reaction may not again manifest itself for several weeks or months. If there is no recurrence of the original complaint it is probable, but not proven, that the toxic effect was caused by a factor other than the drug. In some instances a combination of factors, including the drug, must be present simultaneously for the toxic reaction to occur.

The appearance of toxicity should not in itself be a reason for discarding a new drug since it may be sufficiently valuable therapeutically to justify a measure of risk. For all its potential toxicity, reserpine remains a useful agent for certain patients with essential hypertension. If the physician is careful to warn the patient of the possibility of mental reactions following reserpine administration the danger can be reduced to a minimum. In the cases of many drugs serious toxicity is dose-related. Valuable therapeutic effects can be obtained with little hazard by restricting the upper limit of the dosage range. Digitalis is one example,

as is hydralazine, in which case the lupus syndrome rarely occurs if the daily intake is maintained below 200 mg.

The plan to initiate a central clearing point for the reporting and rapid dissemination of information relating to the toxicity of new drugs appears to be a constructive step. If adequately financed, and if staffed with competent professional personnel, a toxicity data center would perform a valuable function. Such a facility can apply modern data-processing techniques and sophisticated statistical methods to aid in the rapid differentiation between true drug toxicity and sporadic disorders unrelated to the effects of a new agent.

Problems of Dropouts

Long-term evaluation of a new drug is more difficult than the assessment of acute effects. Short-term studies usually permit frequent and close observation of the patient. Medication is dispensed by professional personnel in a hospital or clinic. During long-term evaluation the patient cannot be seen at frequent intervals. Faithful adherence to the prescribed schedule of medication depends on the reliability of the patient whose cooperation may be lost during a prolonged therapeutic trial. Ingestion of medications may then become sporadic or even nonexistent.

The long-term evaluation of a new drug is complicated by dropouts, both the recognized defaulters who fail to return for their appointments and the unknown dropouts who return to the clinic but, for various reasons, no longer take their medication regularly or in prescribed doses. The patients who fail to return for follow-up present a serious problem because the reason for their dropping out usually is unknown. The reason may be ineffectiveness of the drug, occurrence of a toxic reaction, or simply the patient's negligence. The investigator often has no basis for deciding among these various possibilities and, if the dropout rate is large, he will be unable to draw reliable conclusions concerning the therapeutic value or toxicity of the drug under study.

For example, assume that a new drug is being tested in patients with rheumatoid arthritis. Short-term trials suggested that it was effective in controlling acute arthritic symptoms in 75% of the patients treated. This improvement rate is encouraging but not definitive since many therapies, including suggestion, can produce short-lived improvement in this condition. The drug appeared to be nontoxic as judged by laboratory tests during the short-term clinical trial. In the long-term trial there was a 25% dropout rate at the end of six months. In the remaining patients who continued to return for follow-up, an apparent remission was obtained in 50% or slightly more than one third of the total number beginning the trial. Unfavorable reactions occurred in 4% who exhibited febrile reactions with disturbances in liver function tests and in one patient who developed hepatitis and jaundice.

The investigator will be unable to evaluate the drug under such circumstances. If the 25% of patients who defaulted left the clinic because of febrile reactions the drug would be far too toxic to be considered for general use. On the other hand, the dropouts may have undergone a long-term remission and with relief of symptoms felt no need to continue treatment. The investigator would not know which interpretation was correct.

Even more misleading are the cases in which patients maintain an outward appearance of cooperation but actually are not taking the prescribed medication. Such patients are not uncommon—especially in clinic practice. A patient may be collecting disability remuneration or may be obtaining special privileges at his place of employment because of his illness. Some patients return because they can obtain sedatives and hypnotics or analgesics without charge. Alcoholics may alternate between responsible and irresponsible periods and during bouts of drinking may fail to take their medication. Some patients experience uncomfortable symptoms which they ascribe rightly or wrongly to the drug and secretly reduce the dose of the medication below the level at which the symptoms appeared.

The patient who does not take his medication regularly or in prescribed doses will, by so doing, reduce the number of good responders to a test agent. The lack of adequate therapeutic effect will be attributed by the investigator to a failure of the drug. The incidence of toxic effects also will be reduced.

Several techniques can be applied to guard against the problem of dropouts in long-term drug evaluation. One method is to exclude certain patients from the therapeutic trial because of the probability of future default. Alcoholic and psychopathic individuals should be excluded. If a patient is unemployed, although physically able to work, or is constantly changing jobs, he usually is a poor risk. When an individual gives a history of frequent changing of doctors he often will follow the same pattern in the future. The patient should not live far away and the clinic hours must be such that he can conveniently return for regular visits. Admittedly, bias is introduced by such selection. However, the investigator must use discrimination in admitting patients to the study if he is to avoid more serious difficulties later on.

Another helpful technique is to subject each patient to a trial period of several months of standard or placebo treatment before introducing the agent under investigation. During this period many uncooperative individuals drop out before the new drug is introduced.

The patient who fails to take his medicine regu-

larly often can be discovered by the "tablet count" technique. A known number of tablets including a predetermined excess are dispensed and the patient is told to return the bottle at the next visit. A count of the tablets at that time will determine how many doses have been missed. Similarly, the patient who fails to return the bottle or who returns no tablets instead of the predetermined excess discloses a lack of reliability and cooperation.

An additional helpful technique utilizes a harmless marking substance which is incorporated in the tablet. An example of such a substance used in the Veterans Administration Cooperative Study on Antihypertensive Agents is riboflavin. This compound produces fluorescence of the urine under ultraviolet light.

Broadening the Scope in Final Evaluation

It is rarely possible for a single investigator to observe or recognize every toxic reaction or to judge the effectiveness of a new drug under all possible conditions. A given agent may perform well for one investigator and poorly for another. There are various reasons for this: the method of administration and doses may be different; one clinic population may contain more reliable or more responsive patients than another; the responsiveness may depend on sex or age differences in the two clinics, severity of disease, or climatic differences. Even the season of the year influences the response to certain drugs.

The chance distribution of the patients may bias the conclusions of a single investigator. He may discard a useful drug prematurely because of unsatisfactory responses in the initial few patients undergoing treatment. Similarly, if the initial patients in the series respond dramatically well, the investigator may develop a bias favorable to the drug, which impression tends to influence his judgment concerning subsequent failures. Some protection against such bias can be obtained by a double-blind evaluation designed to include enough patients to form a representative sample.

To compensate for unpredictable variations which may lead to an inaccurate and too limited evaluation of a new drug, the scope of the study must be enlarged to include a number of investigators working in different localities. The final evaluation of the usefulness of a new agent, however, will depend on the judgment of clinical practice. A drug which is valuable in the hands of the expert may be too difficult to manage for the physician in general practice. A serious toxic reaction may occur so rarely or in such a limited group of the total population, eg, thalidomide, that it does not become apparent prior to its release for general use. Since such an eventuality usually cannot be anticipated, it is reasonable to consider the first few years following the release of a drug as a continuation of the clinical trial period.

When there is general or nearly unanimous agreement among clinical investigators that a new agent is superior to previously existing compounds, it is difficult to justify further delay in clearance for marketing. If the disease is a serious one the hazard of withholding more effective treatment may be greater than the risk of unrecognized toxicity. However, because the possibility of unsuspected toxicity still exists, any physician in practice observing a severe reaction to a newly released drug should report it promptly to the sponsor or other central agency.

As Modell has stated, "Society must recognize that in its demand for new drugs there is clearly implicit a license for qualified individuals to take certain risks in testing drugs as well as to take calculable risks in using them clinically." The clinical pharmacologist cannot operate effectively if he is subjected to unnecessary restrictions and time consuming or duplicative administrative procedures or is threatened with public censure for unavoidable toxic reactions. New drugs will become scarcer if the additional time, expense, and difficulty involved in obtaining approval will make it unprofitable to produce them.

Summary

Patients undergoing long-term treatment with a new drug must be protected from hazard by prior complete, acute, and chronic trials in animals and acute trials in patients under close observation. Periodic laboratory determinations of renal, hepatic, and bone marrow function and frequent follow-up observation by experienced medical personnel are essential. Prompt reporting of toxicity to the sponsor of the new drug is a responsibility of the entire medical profession and should continue for several years after an agent is released. On the other hand, withdrawal of an effective drug because of questionably related toxicity should not be undertaken prematurely in treatment of a serious disease.

A new therapeutic agent is good or bad in comparison to presently available therapy. The figure of merit for a new drug is based on effectiveness plus ease of administration, minus toxicity, and tolerance. To determine the true value of a new drug studies must be designed which provide an unbiased comparison with placebos or with an established drug or both. The number of defaulters must be held to a minimum and the patients treated should be of sufficient number and variety to form a representative sample of the population suffering from the particular disease being treated.

VA Hospital, Washington, DC.

Generic and Trade Names of Drugs

Reserpine—*Rauloydin, Raurine, Rau-Sed, Reserpoid, Sandril, Serfin, Serpasil, Serpate, Vio-Serpine.*
Rauwolfia serpentina—*Raudixin, Rauserpa, Rauval.*
Hydralazine hydrochloride—*Apresoline Hydrochloride.*

Printed in U.S.A.